Biologically-inspired image processing for machine grasping

Heiko Hoffmann

A primary-visual-cortex-based system allows a robot hand to quickly orient itself and pick up the objects it sees.

Visual processing in mammals is adapted to their behavioral needs: likewise, in visually-guided robots, image processing needs to be suitable for a desired behavior. Thus, the function of the mammalian brain may be a good guideline for choosing the right image-processing techniques for machines. In our work, we make robots learn through experience and thereby study which learning and image-processing techniques lead to a good performance for a given task.

Here, we describe a study in which our goal was to make a robot arm grasp an object presented visually.[1] The robot learned to associate the image of an object with an arm posture suitable for grasping. Learning an association means that there are no world coordinates and there is no tedious calibration of the vision system, instead, the robot learns by randomly exploring different arm postures and by observing the appearance of objects put on a table. Though the emphasis of our work is on learning techniques, here we will focus on the image processing.

We used a robot arm with six joints and a gripper: the vision system was a stereo camera head mounted on a pan-tilt unit (see Figure 1). This setup was located behind a table, which was the operational space and which was visible to the cameras. In training, the robot placed a red brick on the table in random positions and, for each position, recorded an image of the scene after removing the arm. Thus, the training set contains corresponding pairs of grasping postures and object images.

An image can be interpreted as a point in a high-dimensional space (with the number of dimensions equal to the number of pixels). A mapping from such a space to an arm posture suffers from the so-called 'curse of dimensionality ': the distance between pair-wise different images is almost constant, and the orientation of the target gets lost under the dominance of the positional information.[2] Therefore, the image must be pre-processed.
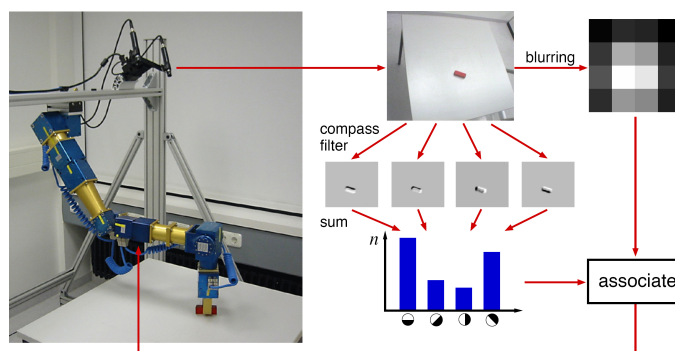

Figure 1. Shown is the information flow in the grasping task. The processing of the camera image is split into two parts. First, to extract position information, the image is blurred and sub-sampled. Second, to extract orientation information, four different compass filters (directional edge filters) extract edges in different directions. The sum of the white pixels in each of the four filtered images results in a histogram of edge distribution. This histogram, together with the blurred image, is associated with an arm posture that enables the robot to grasp the observed object.

The processing technique that was eventually successful was inspired by the function of the visual cortex. The image processing was split into two parts: one for the object 's location and one for its orientation (see Figure 1). To decode the location, the image was first blurred and sub-sampled. Since here the target (the brick) was almost point-like within the camera image, the blurred image is like a population code of the brick 's position. In a population code, many neurons carry information about a parameter: such a code for the retinal location of a stimulus exists also in the primary visual cortex.[3]

To decode the orientation, image filters were used to extract edges in different directions: for each, we counted the edge pixels within the image. This sum was invariant of the brick 's position and was a measure of how close the brick was to a given orientation. Position invariance and orientation tuning are also properties of V1 complex cells.[4]
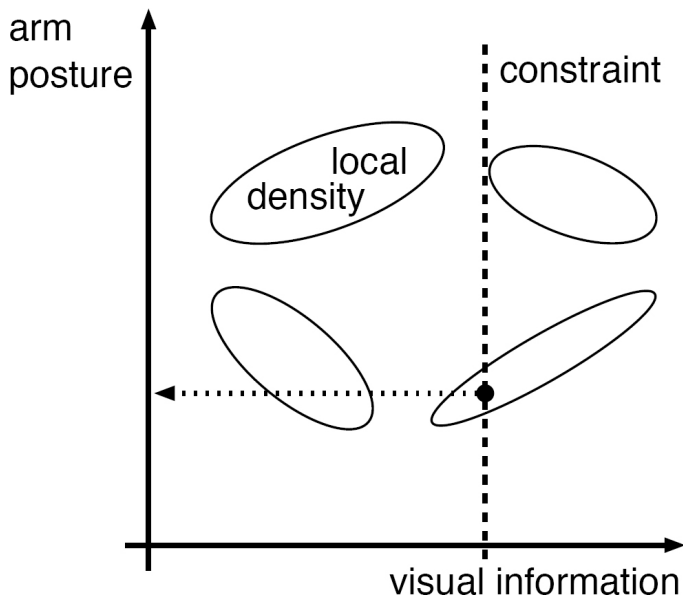
Figure 2. Pattern association. Training patterns lie in the product space of arm posture and visual information. The density of the pattern 's distribution is modeled by a mixture of Gaussian functions (ellipses are iso-density curves). To map the visual information onto an arm posture, we define the output space as a constrained space anchored at the input. On this subspace, the highest local density gives the desired output.

The resulting visual information could be used to first learn and then to recall the association with an appropriate arm posture for grasping (Figure 2). Specifically, the decomposition of the image processing into two parts and the use of population codes kept the grasping errors low.[1,2] This robot experiment demonstrated that brain functions can provide guidelines for robotic control, but also robots can help us to understand the brain. This is done by first demonstrating that certain (often hypothetical) functions actually work and then showing the advantages of certain data-processing techniques in a behavioral context.

Author Information

Heiko Hoffmann
Max Planck Institute for Human Cognitive and Brain Sciences
Munich, Germany
http://www.heikohoffmann.de/publications.html

References
1. H. Hoffmann, W. Schenck, and R. Möller, *Learning visuomotor transformations for gaze-control and grasping*, **Biological Cybernetics 93**, pp. 119–130, 2005.
2. H. Hoffmann, **Unsupervised Learning of Visuomotor Associations**, **11 MPI Series in Biological Cybernetics**, Logos Verlag, Berlin, 2005.
3. D. Jancke, W. Erlhagen, H. R. Dinse, A. C. Akhavan, M. Giese, A. Steinhage, and G. Schöner, *Parametric population representation of retinal location: Neuronal interaction dynamics in cat primary visual cortex*, **J. of Neuroscience 19** (20), pp. 9016–9028, 1999.
4. T. W. Kjaer, T. J. Gawne, J. A. Hertz, and B. J. Richmond, *Insensitivity of V1 complex cell responses to small shifts in the retinal image of complex patterns*, **J. of Neurophysiology 78**, pp. 3187–3197, 1997.